

# EXHIBIT D

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 1

UNITED STATES DISTRICT COURT  
NORTHERN DISTRICT OF CALIFORNIA  
SAN FRANCISCO DIVISION

Richard Kadrey, et al.,	)	
Individual and Representative	)	
Plaintiffs,	)	CASE NO.
	)	3:23-cv-03417-VC
-against-	)	
	)	
Meta Platforms, Inc.,	)	
Defendant.	)	
	)	

\*\*\* HIGHLY CONFIDENTIAL \*\*\*

ATTORNEYS' EYES ONLY  
VIDEO-RECORDED DEPOSITION OF  
MELANIE KAMBADUR

Cooley, LLP  
55 Hudson Yards  
New York, New York 10001

09/17/2024  
9:07 a.m. (EDT)

REPORTED BY: MONIQUE CABRERA

DIGITAL EVIDENCE GROUP  
1730 M Street, NW, Suite 812  
Washington, D.C. 20036  
(202) 232-0646

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 66

1 with Common Crawl in order to use the Common  
2 Crawl data sources for the purposes of  
3 pre-training?

4 MR. WEINSTEIN: Object to form.

5 A. I believe there might be a license  
6 attached to Common Crawl.

7 BY MR. YOUNG:

8 Q. What about the works inside the  
9 Common Crawl data sources? Do you know if there  
10 is license attached to those?

11 A. It's possible for some of them.

12 Q. Do you know if Common Crawl licensed  
13 with any of the works that it scraped in order to  
14 include it in the data source?

15 MR. WEINSTEIN: Object to form.

16 A. I do not know.

17 BY MR. YOUNG:

18 Q. Do you know what a shadow library  
19 is?

20 A. I don't recall the term.

21 Q. Do you know what Library Genesis is?

22 A. At a high level, yes.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 67

1 Q. Is it also known as Libgen?

2 A. Yes.

3 Q. What do you understand that to be?

4 A. A dataset that contains books and  
5 scientific articles.

6 Q. Was Libgen used to pre-train any of  
7 the Llama models?

8 A. Yes.

9 Q. Which ones?

10 A. I know of it being used for Llama 3.  
11 It was not used for Llama 1 and Llama 2.

12 Q. Llama 3.1 is from the paper. Is  
13 Llama 3 different from Llama 3.1?

14 A. Llama 3.1, yes, it's different.

15 Q. Was Libgen used to pre-train the  
16 Llama 3 series of models?

17 A. Yes.

18 Q. So you mentioned that you had a high  
19 level of understanding of what Libgen was. Do  
20 you know specifically what type of media are  
21 included in Library Genesis?

22 A. No, I don't know specifically.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 68

1 Q. But you know it contains books?

2 A. Yes.

3 Q. Do you know if it contains  
4 copyrighted books?

5 A. I don't know the specific books on  
6 Libgen.

7 Q. Have you ever examined what was in  
8 the Libgen dataset?

9 MR. WEINSTEIN: Object to form.

10 A. I recall seeing some high level  
11 summary statistics.

12 BY MR. YOUNG:

13 Q. Do you know if there are any books  
14 that are written in the last 50 years on Libgen?

15 A. I believe so. But, again, I haven't  
16 directly seen -- I don't directly recall any  
17 books titles.

18 Q. Was there someone at Meta who was  
19 responsible for the implementation of Libgen as  
20 pre-training data for the Llama 3 models?

21 A. What do you mean "the implementation  
22 of Libgen"?

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 69

1 Q. Was someone responsible for curating  
2 Libgen for use as pre-training data?

3 A. Yes.

4 Q. Who was that person?

5 A. Nikolay Bashlykov.

6 Q. And he lives in London?

7 A. I'm not actually sure where he  
8 lives. He works from the London office.

9 Q. Do you know how Mr. Bashlykov  
10 obtained the Libgen dataset?

11 A. No.

12 Q. Do you understand Mr. Bashlykov to  
13 be the source of the Libgen dataset for Meta's  
14 use as pre-training data?

15 MR. WEINSTEIN: Object to form.

16 A. I don't recall.

17 BY MR. YOUNG:

18 Q. But Mr. Bashlykov had primary  
19 responsibility for implementing Libgen or  
20 including Libgen into the Llama 3 models.

21 Would that be fair to say?

22 A. I don't know if I would say it that.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 70

1 We said he worked on pre-processing the data,  
2 that is fair to say.

3 Q. What is pre-process -- processing?  
4 Excuse me.

5 A. That means transforming the data  
6 with a series of steps that make it more suitable  
7 for training models.

8 Q. Does that include removing -- strike  
9 that.

10 What kind of steps would be involved  
11 in pre-processing?

12 A. Are you asking specifically for  
13 Llama models or more generally?

14 Q. Yeah, for the whole Llama models.

15 A. I don't know all the steps. But for  
16 example, we perform various types of  
17 de-duplication. We tokenize the data, which  
18 means -- yeah, we tokenize the data. We chunk  
19 the data into spans of text.

20 Q. Let's talk about de-duplication. So  
21 if the models -- if a model is trained or  
22 pre-trained on the same text over and over again,

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 71

1 would that affect the model's performance?

2 A. It could.

3 Q. How could it affect the model's  
4 performance?

5 A. It could make the training less  
6 efficient to repeatedly seem -- see the same  
7 data.

8 Q. So you want to limit the model's  
9 exposure to common text?

10 Would that be fair to say?

11 A. What do you mean by "common text"?

12 Q. For example -- so de-duplication,  
13 right, you would want to remove text that appears  
14 frequently from the pre-training dataset.

15 Is that -- would that be fair to  
16 say?

17 A. It depends on the text.

18 Q. So, for example, if you -- you know,  
19 you -- you've read books, correct?

20 A. Yes, I've read books.

21 Q. Okay. So you know in the very front  
22 of the page, there is the page including all the



9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 72

1 copyright information? For example, "published  
2 by," all the have information.

3 A. Yes.

4 Q. Would that be the type of  
5 information you would remove before processing?

6 A. Potentially.

7 Q. Was that information actually  
8 removed for data used to pre-train the Llama  
9 models?

10 A. I don't recall.

11 Q. Do you know -- going back to Libgen.  
12 Do you know if it contained fiction  
13 books?

14 A. I believe it did.

15 Q. How about non-fiction books?

16 A. Yes, I believe it did.

17 Q. Are you familiar with the Project  
18 Gutenberg?

19 A. At a very high level, yes.

20 Q. What do you understand Project  
21 Gutenberg to be?

22 A. A -- I -- I don't know the full

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 315

1 MR. WEINSTEIN: Object to form.

2 A. Sorry. Your -- your question is not  
3 matching what Nisha said, so I don't know how to  
4 answer that.

5 BY MR. YOUNG:

6 Q. Okay. Well, you respond to Nisha,  
7 right, at 14:24:13?

8 A. Yes.

9 Q. You say: I feel like that's on the  
10 people who distributed the dataset somewhat, no?

11 Right?

12 A. Yes.

13 Q. So would -- were you laying blame on  
14 the EleutherAI if -- if the Pile contained  
15 copyrighted work?

16 A. I don't know that I would -- yeah.  
17 I don't know that I can make that statement. I  
18 don't recall exactly what I meant.

19 Q. So the "people" -- the "people" in  
20 the sentence, do you think that refers to  
21 EleutherAI?

22 A. I suppose.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 316

1 Q. And then 14:31:58, do you see that  
2 sentence there beginning with "um"?

3 A. Yes.

4 Q. Okay. And then you say: To be  
5 honest, sometimes we do move to copyright text as  
6 part of the common cleaning because it is -- it  
7 is very repeated across datasets.

8 Do you see that?

9 A. Yes.

10 Q. Okay. Do you recall us test --  
11 discussing this very early in this deposition  
12 about removing kind of the front pages of the  
13 books?

14 A. I recall discussing this area.

15 Q. All right. So does that mean you --  
16 you removed copyrighted text as part of  
17 pre-processing of pre-training data?

18 MR. WEINSTEIN: Object to form.

19 A. It sounds like, at this point in  
20 time, I had an understanding that we at least  
21 move it and possibly that we could de-duplicate  
22 it.

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 317

1 MR. YOUNG: Okay. I'm done with  
2 this document and I've got one more line of  
3 questioning. So about -- maybe about an half  
4 hour left. So I'll leave it to you guys.

5 Do you guys want to take a short  
6 break or do you guys want to keep trekking?

7 THE WITNESS: I'm either way.

8 MR. WEINSTEIN: We can proceed, I  
9 think.

10 MR. YOUNG: Okay. Great.

11 And then we are going to mark as  
12 Plaintiffs' 22, a document bearing Bates  
13 label Meta Kadrey 00054518.

14 (Whereupon, Plaintiffs' Exhibit 22,  
15 Bates Number Meta Kadrey 00054518, was marked  
16 for identification.)

17 BY MR. YOUNG:

18 Q. So take a couple moments to review  
19 the document and let me know when you're ready.

20 A. I reviewed the document.

21 Q. All right. And is Plaintiffs' 22 a  
22 Workplace Chat between you and yourself -- you

9/17/2024

Richard Kadrey, et al. v. Meta Platforms, Inc.  
Highly Confidential - Attorneys' Eyes Only

Melanie Kambadur

Page 337

1 CERTIFICATE OF SHORTHAND REPORTER NOTARY PUBLIC

2 I, Monique Cabrera, the officer  
3 before whom the foregoing deposition was  
4 taken, do hereby certify that the foregoing  
5 transcript is a true and correct record of  
6 the testimony given; that said testimony was  
7 taken by me stenographically and thereafter  
8 reduced to typewriting under my direction;  
9 and that I am neither counsel for, related  
10 to, nor employed by any of the parties to  
11 this case and have no interest, financial or  
12 otherwise, in its outcome.

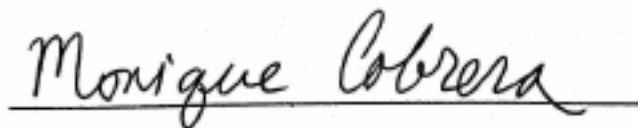
13 IN WITNESS WHEREOF, I have hereunto  
14 set my hand this 17th day of September, 2024.

15

16

17

18

A handwritten signature in cursive script that reads "Monique Cabrera". The signature is written in dark ink and is positioned above a horizontal line.

19

MONIQUE CABRERA

20

Notary Public in and for the State of New York

21

County of Suffolk

My Commission No.

22

Expires: 06/12/2026